

Interviewer effects when investigating abuse were not compatible with effect modification but instead with confounding

Sílvia Fraga^{a,b,*}, Raquel Lucas^{a,b}, Diogo Costa^{a,b}, Henrique Barros^{a,b}

^aDepartment of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School, Al. Prof. Hernâni Monteiro, 4200-319 Porto, Portugal

^bInstitute of Public Health, University of Porto (ISPUP), Rua das Taipas, 135-139, 4050-600 Porto, Portugal

Accepted 24 July 2012; Published online 4 May 2013

Abstract

Objectives: To describe interviewer-related variability in abuse estimates and assess the nature of the interviewer effects on the associations between elder abuse and covariates.

Study Design and Setting: After intensive training, six interviewers administered structured questionnaires through face-to-face interviews to assess abuse in a population-based sample of 641 Portuguese individuals aged 60–84 years.

Results: The overall prevalence of abuse victimization during the previous year was 28.1%, but it differed significantly according to the interviewer, ranging from 16.9% to 36.8%. There was no statistical effect modification introduced by the interviewer on the association of abuse and its determinants. Additionally, interviewer-level variables (empathy and violence beliefs) showed no significant contribution to explain the variance attributable to potential interviewer effects. Adjusting for the interviewer had little or no effect on the odds ratio of abuse for gender, age, education, and quality of life. However, the interviewer introduced relevant confounding of the associations between abuse and other sensitive topics, such as somatic complaints.

Conclusion: Although no relevant effect modification was observed, this study emphasizes the importance of the interviewer as a relevant confounder when estimating associations between sensitive variables, as it is the case of elder abuse. © 2013 Elsevier Inc. All rights reserved.

Keywords: Bias; Epidemiologic; Interviewer effects; Effect modifier; Confounding; Elder abuse

1. Introduction

Interpersonal violence is a violation of human rights and a major public health concern [1]. However, controversy in the conceptualization, definition, and measurement of abuse increases the difficulty in ascertaining the frequency and patterns of its different forms [2]. The most common way to assess the history of abuse is to directly prompt participants about their experiences and behaviors through the interview method.

Face-to-face interviewing has various advantages compared with other methods: the identity of the respondent can be ascertained by the interviewer, it decreases missing

items in the questionnaire, and it provides an environment that may help overcome communication barriers. However, it is more expensive and may carry an interviewer effects arising from the close interaction between the interviewer and the interviewee [3–5].

The debate over interview accuracy remains, and a long tradition in proposing and identifying interviewer effects can be traced back as early as the first surveys were designed to measure the health status of populations [6]. Interviews might thus be a major source of bias in epidemiologic research, and a large bulk of literature has been devoted to this debate and to the development of strategies to minimize it [4–6].

The use of standardized questionnaires to induce equivalent item phrasing and an interview orientation protocol are recommended to effectively minimize the expected misclassification [4]. Additionally, the training of interviewers should provide a common questioning frame and similar strategies to handle unusual or unexpected circumstances during the interviewing process. It is common to incorporate quality control measures into the design of

Conflict of interest: The authors declared no conflict of interests.

Funding: This work was supported by the Executive Agency for Health and Consumers (Project A/100929) and Portuguese Foundation for Science and Technology (SFRH/BD/44408/2008) to S.F., (SFRH/BD/40656/2007) to R.L., and (SFRH/BD/66388/2009) to D.C. and (PTDC/SAU-SAP/122904/2010).

* Corresponding author. Tel.: +351-222061820; fax: +351-222061821.

E-mail address: silfraga@med.up.pt (S. Fraga).

What is new?

- The interviewer does not seem to be a major effect measure modifier of the estimates of associations between elder abuse and covariates.
- The study emphasizes the importance of the interviewer as a confounder when estimating associations between sensitive variables.
- Adjusting for the interviewer variable in data analysis is advisable when estimating associations between sensitive variables.

epidemiologic studies to minimize interviewer effects, but few researchers report which measures they use, examine the data for interviewer variation, or explore the impact of such variation on study findings [7].

Even if standard strategies to minimize interviewer effects are incorporated into the study protocol studies addressing sensitive topics, such as those that concern intimate personal behaviors, such as history of abuse, may remain specially prone to interviewer effects [8]. These effects may be attributable to the characteristics of the interviewer or the respondent and interactions between them.

Particular characteristics of participants, such as age-related cognitive decline, pose difficulties to disclosure and hence, to research, worsened when dealing with sensitive topics as violence, in which there may be substantial interaction with the interviewer's characteristics [9]. Behind attitudes and characteristics of the interviewer and respondents, the context in which they live cannot be neglected.

Another important issue studied in the last decades has been gender effect. Although some studies show differences in results when interviews were performed by female or male interviewers, it is still unclear whether there are gender differences in the validity of data collected and if or when interviewers and respondents should be matched by gender [8].

Beyond patent characteristics of interviewer or interviewee such as demographic traits, latent aspects may influence reporting of abuse. In particular, attributes of the interviewer, such as personality traits, attitudes, or a personal experience with abuse may affect the response [4]. Also, sensitivity to violence may also condition the way the information was collected as well its disclosure [10].

In the presence of interviewer effects, that is, a misclassification bias in any outcome measured, researchers need to characterize the nature direction and extent of this influence on their estimates. In epidemiologic terminology, this translates into assessing whether the interviewer variable should be dealt with as an effect measure modifier,

a confounder, or none of these. The investigation of this issue has a relevant impact on how to conduct data analysis, particularly regarding the need for stratification or adjustment.

The objectives of this study were to describe interviewer-related variability in abuse estimates and assess the nature of the interviewer effects on the associations between elder abuse and covariates.

2. Participants and methods

2.1. Participants

The present research used the Portuguese sample of participants in the international collaboration named by the acronym ABUEL, a large study involving universities of seven countries and cofunded by the European Commission. The study design and sample have been fully described elsewhere [11,12].

In brief, Portuguese participants were urban dwellers previously recruited as part of a population-based cohort of adults living in Porto, Portugal (the EPIPorto study). The participant selection was conducted during 1999–2003 using random digit dialing. Households were the sampling frame, followed by simple random sampling to select one eligible person among permanent residents in each household [13].

By 2009, 845 subjects within the original EPIPorto cohort met the age criteria (≥ 60 years) for the ABUEL study, and they were contacted to participate in the present study. However, 65 individuals could not be reached, 83 refused to participate, 28 were deceased, and 2 had missing information on the questionnaire. Of the 667 individuals who accepted to participate, 11 were excluded from the analysis because of significant cognitive impairment (Folstein's Mini-Mental State Examination score < 24). We have excluded 15 participants from the present investigation who self-completed the questionnaires. The final sample comprised 641 participants.

No statistically significant differences were observed between participants and nonparticipants regarding baseline characteristics such as gender, age, education, marital status, smoking, and alcohol drinking. The local ethics committee (Hospital São João) approved the ABUEL study protocol.

2.2. Interviewers

From a pool of 30 candidates, we selected six female interviewers based on their professional background, experience with research projects, and previous work with the elderly. They were aged 25–30 years and possessed a degree in social sciences.

Intensive training was followed during a week. Interviewers were introduced to the study protocol, which comprised detailed information on interview procedures and special care for confidentiality and respect over the

participant's will. Training also comprised an in-depth discussion of the questionnaire and exhaustive interpretative possibilities of questions, role playing of specific probable situations, and several reminders for correct assessment of sensitive topics, ensuring ethical and safe disclosure of information. Also, interviewers were closely monitored and trained throughout the study. For the purpose of the present analysis, interviewers were labeled A through F.

2.3. Procedures

A letter of introduction giving notice of the study was sent by post to all selected individuals. Following the letter, the individuals were contacted by telephone, and those who were willing to take part were scheduled for interview. Up to four attempts were made to contact each selected participant by telephone. When individuals refused to participate, no further contact attempts were made. The timing of calls varied to maximize the chances of contact. Calls were conducted at different times of the day during weekdays, with subsequent calls in the evening if no answer was obtained during the day and finally with calls on weekends if no answer was obtained during the week. If no contact was established by telephone, participants were deemed as noncontacted.

On the appointment day, interviewers recalled participants of the study objectives and clarified any doubts. Then, a consent form was signed. The study-structured questionnaire (available at <http://www.abuel.org>) was administered face to face. The duration of the interview was recorded, and working shifts were assigned so that all interviewers conducted a similar number of interviews at each period of the day. To ensure a balanced distribution of participants between interviewers, interviewer assignments were made on a consecutive basis, regardless of participants' characteristics.

Interviewer F dropped out from the study before completion, resulting in the redistribution of the previously assigned participants to other interviewers and in a smaller number of participants evaluated by that specific interviewer. Questionnaires assessed information on social and health characteristics. Lifestyle variables measured were alcohol drinking and smoking.

Depressive and anxiety symptoms were measured with the Hospital Anxiety and Depression Scale [14] consisting of 14 items (graded 0–3), seven each about depression and anxiety. High scores correspond to high depression and anxiety levels. Somatic complaints were measured with the short version of the Giessen Complaint List consisting of 24 items (graded 0–4) about various somatic symptoms. High scores correspond to high levels of somatic complaints.

Quality of life (QoL) was measured with the World Health Organization Quality of Life Assessment for Older Adults [15] consisting of 24 items (graded 1–5). The total score amounts to 100 and high scores correspond to high QoL (total/subscales).

Elder abuse was considered when there was self-reported physical, psychological, financial, and sexual abuse against an individual aged over 60 years.

Psychological abuse was defined as the infliction of mental anguish and it was assessed using 11 items (e.g., insults), physical abuse was assessed by 17 items that described the infliction of pain or injury (e.g., beatings), sexual abuse was assessed through eight items and it includes nonconsensual contact of any kind with an older person (e.g., intercourse against one's will), and financial abuse was measured through nine items related to the illegal or improper exploitation and/or use of funds or resources (e.g., forcibly taken money).

For each type of abuse, participants were asked to state its frequency, considering the following categories: never, 1 time; 2 times; 3–5 times; 6–10 times; 11–20 times; and more than 20 times. To define a dichotomized variable for each type of abuse, we have collapsed items of each one. Abuse in the previous 12 months was considered present if the participant reported any type of abuse (physical, psychological, sexual, or financial) in the past year and absent if they reported abuse only before the previous year or if it had never happened.

Personality characteristics and attitudes of interviewers were measured after data collection was finished using two self-administered scales: the Interpersonal Reactivity Index (IRI) [16] and the beliefs about marital abuse scale, a specific Portuguese scale about violence tolerance in the Portuguese population [Escala de Crenças sobre Violência Conjugal (ECVC)] [17].

The IRI aims to assess the global concept of empathy, and it was translated to Portuguese for the aim of the present work [18]. It is a 21 items scale answered in a five-point Likert scale ranging from 1 (does not describe me well) to 5 (describes me very well). Higher score means higher empathy level.

The ECVC has 25 considering common beliefs regarding violence, in which the interviewers should be positioned in terms of agreement or disagreement. Higher scores mean higher tolerance toward violence.

The five interviewers who remained in the study until the end completed both scales. We hypothesized that, if the interviewer effects were because of these characteristics measured by the two scales, we would find a direct relation between empathy score and abuse and an inverse relation between ECVC (tolerance to marital violence) score and elder abuse.

2.4. Data analysis

Chi-square test or Fisher's exact test was used to examine differences with categorical variables, and continuous variables were compared between groups using analysis of variance or Kruskal–Wallis tests. Logistic regression was used to compute odds ratios (ORs) and 95% confidence intervals (95% CIs) for the association of abuse and its covariates.

Questionnaires assessed information on social and health characteristics. Lifestyle variables measured were alcohol drinking (grouped as more than three times vs. three times or less per week) and smoking (current smokers vs. never- or ex-smokers).

To perform logistic regression, we dichotomized continuous covariates such as age (≤ 70 vs. > 70 years), education (≤ 4 vs. > 4 years), somatic complaints (scoring ≤ 19 vs. > 19), anxiety (scoring ≤ 5 vs. > 5), and QoL (≤ 89 vs. > 89), taking the median values as cutoffs. Likelihood ratio test was reported to test the interaction term (interviewer \times independent variable) for the associations.

Taking into account that a two-level hierarchical structure should provide a more accurate description of the underlying structure of data, we used multilevel linear regression models, considering the participant as the first level and the interviewer as the second level. We built three models to account for different level variables. Model 0 (crude) analyzed the abuse variance by interviewer. Model 1 was added of the service variables identified to be associated with the outcome, that is, gender, age, education, anxiety, somatic complaints, and QoL, allowing to understand the effect of individual participants' characteristics in the differences between interviewers. Model 2 was added to the individual-level variable, interviewer-level characteristics such as tolerance toward abuse and empathy. These effects were measured by proportional change in variance from model 0, and intraclass correlation coefficients (ICCs) were computed to show the percentage of observed variation in the abuse attributable to interviewer-level characteristics.

3. Results

3.1. Frequency of elder abuse by interviewer

The overall prevalence of abuse victimization during the previous year was 28.1%, and it was significantly different between interviewers, 16.9% being the lowest and 36.8% being the highest computed prevalence estimates ($P = 0.021$) (Table 1). The mean (standard deviation) duration of the interview was significantly different according to interviewers, ranging from 89 (23.8) to 119 (45.4) minutes ($P < 0.001$). Additionally, significant differences between interviewers in median scores (25th–75th percentile) were observed for somatic complaints, anxiety, and regarding QoL, from 14.0 (6.0–24.0) to 27.0 (17.0–54.20), from 4.0 (2.0–8.0) to 7.0 (4.0–10.0), and from 85.0 (77.0–92.2) to 90.0 (82.0–98.0) (Table 2), respectively.

The social and demographic characteristics of participants (gender, age, marital status, living situation, and education) were similarly distributed between the interviewers and the distribution of behavioral and lifestyle characteristics of participants, such as smoking, drinking alcohol, and health care utilization (Table 2).

3.2. Assessment of effect measure modification and confounding by interviewer

The magnitude and direction of the associations between abuse and covariates, overall and according to interviewer, are presented in Table 2. No statistically significant interaction between interviewer and assessed exposures was found. However, regarding the association for abuse and sociodemographic covariates, our results indicate that males presented lower odds of being victims (OR, 0.75; 95% CI: 0.52, 1.07), and these estimates ranged from OR = 0.52 to OR = 1.16 between interviewers. Regarding age, the estimate was OR = 0.94 (95% CI: 0.66, 1.32) for the comparison between older age group (> 70 years old) and youngest (≤ 70 years old), ranging from OR = 0.57 to OR = 2.15 by interviewer. For higher compared with lower educational level, the overall abuse point estimate was OR = 0.76 (95% CI: 0.54, 1.08), ranging from OR = 0.40 to OR = 1.62 between interviewers. Odds for abuse victimization were higher for those presenting somatic complaints compared with those who did not present such complaints (OR, 1.49; 95% CI: 1.05, 2.10), with results ranging from OR = 0.93 to OR = 2.46, and for anxiety symptoms, the estimate was OR = 1.86 (95% CI: 1.31, 2.63), ranging from OR = 0.95 to OR = 3.33 between interviewers. Significantly lower odds of scoring high in QoL were observed for those stating exposure to abuse OR = 0.39 (95% CI: 0.27, 0.57) with estimates varying from OR = 0.26 to OR = 1.39 between interviewers (Table 2).

Adjusting for the interviewer had little or no effect on the OR of abuse for covariates such as gender, age, education, and QoL. However, it had a more relevant effect on the OR for somatic complaints (crude estimate OR, 1.49 and adjusted estimate OR, 1.35) and anxiety (crude estimate OR, 1.86 and adjusted estimate OR, 1.77).

3.3. Hierarchical modeling of interviewer effects

In the scale of violence beliefs (score ranging from 25 to 125), the mean was 41.8 (14.48). In the IRI, results showed a mean of 58.6 (6.80) in the empathy scale. Results from the fully adjusted random effects model (Table 3, model 2) showed that all estimates remained similar for the association of abuse with individual-level and interviewer-level considered factors. QoL score was the only statistically significant predictor (OR, 0.39). A drop of 3.6% in the ICC was observable from the crude (21.4%) to the fully adjusted random effects model (17.8%), expressing a small effect of the interviewer's characteristics on the remaining variance explained by second-level variables. Therefore, the interviewer-level variables considered (empathy and violence beliefs) did not seem to be significant contributors for explaining the variance attributable to potential interviewer effects.

Table 1. Previous year prevalence of self-reported elder abuse, sociodemographic, health, and behavioral participants' characteristics according to interviewers

Characteristics of participants and interviews	Interviewer code						P
	A	B	C	D	E	F	
Number of interviews	106	130	112	125	111	57	
Psychological abuse, n (%)	30 (28.3)	16 (12.3)	25 (22.3)	31 (24.8)	25 (22.5)	15 (26.3)	
Physical abuse, n (%)	4 (3.8)	2 (1.5)	1 (0.9)	3 (2.4)	2 (1.8)	2 (3.5)	
Sexual abuse, n (%)	0	2 (1.5)	1 (0.9)	2 (1.6)	0	3 (5.3)	
Financial abuse, n (%)	10 (9.4)	9 (6.9)	8 (7.1)	8 (6.4)	14 (12.6)	2 (3.5)	
Any type of abuse, n (%)	39 (36.8)	22 (16.9)	29 (25.9)	38 (30.4)	34 (30.6)	18 (31.6)	*
Gender, n (%)							
Female	62 (58.5)	72 (55.4)	67 (59.8)	82 (65.6)	68 (61.3)	38 (66.7)	
Male	44 (41.5)	58 (44.6)	45 (40.2)	43 (34.4)	43 (38.7)	19 (33.3)	
Age, years, mean \pm SD	70.7 \pm 6.8	71.7 \pm 6.7	70.3 \pm 6.7	69.7 \pm 6.3	70.2 \pm 6.7	69.4 \pm 7.5	
Marital status, n (%)							
Single	2 (1.9)	11 (8.5)	4 (3.6)	6 (4.8)	5 (4.5)	6 (10.5)	
Married/cohabit	74 (69.8)	80 (61.5)	79 (70.5)	82 (65.6)	65 (58.6)	33 (57.9)	
Divorced/separated	8 (7.5)	15 (11.5)	4 (3.6)	11 (8.8)	9 (8.1)	3 (5.3)	
Widowed	22 (20.8)	24 (18.5)	25 (22.3)	26 (20.8)	32 (28.8)	15 (26.3)	
Education, n (%)							
Less than primary	13 (12.3)	11 (8.5)	5 (4.5)	21 (16.8)	18 (16.2)	5 (8.8)	
Primary	37 (34.9)	48 (36.9)	41 (36.6)	42 (33.6)	45 (40.5)	21 (36.8)	
Secondary	41 (38.7)	56 (43.1)	45 (40.2)	37 (29.6)	32 (28.8)	22 (38.6)	
University	15 (14.2)	15 (11.5)	21 (18.8)	25 (20.0)	16 (14.4)	9 (15.8)	
Living alone, n (%)	16 (15.1)	32 (24.6)	21 (18.8)	26 (20.8)	30 (27.0)	12 (21.1)	
Current smoker, n (%)	4 (3.8)	8 (6.2)	7 (6.3)	7 (5.6)	8 (7.2)	5 (8.8)	
Drinking more than three times/week, n (%)	56 (52.8)	69 (53.1)	52 (46.4)	56 (44.8)	52 (46.8)	24 (42.1)	
Health care utilization, n (%)	106 (100.0)	128 (98.5)	109 (97.3)	122 (97.6)	105 (94.6)	56 (98.2)	
Length of interview, minutes, mean (SD)	113.06 \pm 23.81	88.89 \pm 23.79	92.34 \pm 20.57	92.33 \pm 24.85	118.93 \pm 45.40	115.51 \pm 28.24	***
Somatic complaints, median (P25-P75) ^a	22.0 (9.7–32.2)	14.0 (6.0–24.0)	17.0 (11.0–26.0)	21.0 (9.5–33.5)	19.0 (11.0–29.0)	27.0 (17.0–54.20)	***
Depression, median (P25-P75) ^b	5.0 (3.0–9.0)	4.0 (2.0–6.2)	5.0 (3.0–8.0)	5.0 (3.0–8.0)	6.0 (2.0–9.0)	5.0 (3.0–8.5)	***
Anxiety, median (P25-P75) ^b	6.0 (4.0–10.0)	4.0 (2.0–8.0)	5.0 (3.0–7.0)	5.0 (2.0–7.5)	4.0 (3.0–7.0)	7.0 (4.0–10.0)	***
Quality of life, median (P25-P75) ^c	85.0 (77.0–92.2)	90.0 (82.0–98.0)	89.5 (80.0–97.0)	88.0 (79.0–95.0)	90.0 (82.0–98.5)	87.0 (79.5–90.0)	*

Abbreviations: SD, standard deviation; HADS, Hospital Anxiety and Depression Scale.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ for comparison between interviewers.

^a Giessen Complaint List.

^b HADS.

^c World Health Organization Quality of Life-old.

Table 2. Odds ratio (OR) and 95% confidence intervals (95% CIs) between self-reported elder abuse and gender, age, education, somatic complaints, anxiety, and quality of life according to interviewer

Characteristics of participants	Overall	Adjusted for interviewer	Stratified by interviewer code					
			A	B	C	D	E	F
Gender (male)	0.75 (0.52, 1.07)	0.76 (0.53, 1.10)	0.69 (0.31, 1.56)	0.52 (0.20, 1.38)	1.07 (0.45, 2.53)	0.58 (0.25, 1.36)	1.16 (0.51, 2.64)	0.69 (0.20, 2.33)
Age (> 70 years)	0.94 (0.66, 1.32)	0.98 (0.69, 1.39)	0.77 (0.35, 1.70)	1.25 (0.48, 3.23)	0.57 (0.24, 1.34)	0.87 (0.40, 1.86)	2.15 (0.94, 4.92)	0.82 (0.26, 2.57)
Education (> 4 years)	0.76 (0.54, 1.08)	0.77 (0.54, 1.10)	0.40 (0.18, 0.89)	0.80 (0.32, 2.00)	0.67 (0.29, 1.58)	1.62 (0.75, 3.49)	0.88 (0.39, 2.01)	0.56 (0.18, 1.72)
Somatic complaints ^a (scoring > 19)	1.49 (1.05, 2.10)	1.35 (0.95, 1.93)	2.46 (1.07, 5.65)	0.93 (0.35, 2.49)	1.49 (0.63, 3.49)	1.21 (0.56, 2.59)	1.03 (0.46, 2.30)	1.30 (0.38, 4.44)
Anxiety ^b (scoring > 5)	1.86 (1.31, 2.63)	1.77 (1.24, 2.54)	2.06 (0.88, 4.82)	2.37 (0.93, 6.00)	3.33 (1.38, 8.04)	1.25 (0.57, 2.72)	0.95 (0.41, 2.22)	2.01 (0.60, 6.74)
Quality of life ^c (scoring > 89)	0.39 (0.27, 0.57)	0.41 (0.28, 0.61)	0.31 (0.13, 0.75)	0.26 (0.10, 0.71)	0.26 (0.10, 0.66)	0.53 (0.23, 1.21)	0.58 (0.22, 1.51)	1.39 (0.40, 4.91)

Abbreviation: HADS, Hospital Anxiety and Depression Scale.

Likelihood ratio test for the interaction term was used for all exposures, all of them presented a *P*-value > 0.05.^a Giessen Complaint List.^b HADS.^c World Health Organization Quality of Life-old.

4. Discussion

The overall prevalence of abuse victimization during the previous year was significantly different according to the interviewer, ranging from 16.9% to 36.8%.

Considering each type of abuse, differences were most pronounced for psychological violence, which was also the most commonly reported type of abuse. As a consequence, a similar variation is observable when looking at the aggregated abuse variable. Because other types of abuse were much less frequent, our overall abuse frequency is to a higher extent a measure of psychological violence.

Evidence that there is heterogeneity when the most subjective or sensitive issues are measured was also supported by the variation in the distribution of somatic complaints, anxiety, and QoL. A possible explanation for such differences is heterogeneity in data collection procedures. However, measures commonly used to minimize interviewer variation were incorporated into the study design. Namely, to minimize interviewer bias, a highly structured questionnaire was used. Additionally, interviewers were introduced to the protocol that intended to provide standardized ways of reformulating a question if it was not initially understood and procedures rules about how to interpret ambiguous answers. Still, it may be possible that the interviewer differences detected were due in part to the allocation process of participants by interviewer. In the present study, this explanation seems unlikely because responses to questions on sociodemographic and behaviors were similarly distributed among interviewers.

The remaining factors most likely to explain such variation in estimates are interviewer effects, which result from a complex set of sources of variability. It has been described that female interviewers have fewer refusals and higher completion rates than male [19,20]. They may create conditions more conducive to disclosure and be perceived as more sympathetic than male interviewers [21]. In our study, to avoid this gender effect, we selected only female interviewers. We also chose interviewers with similar professional background and in the same age range because these sociodemographic factors could affect the magnitude of information bias.

With other factors excluded, we expected that the interviewer variation could be related to empathy or attitudes toward violence from the interviewers, shown as a direct relation between empathy score and abuse and as an inverse relation between ECVC (tolerance to marital violence) score and elder abuse. However, these characteristics did not seem to explain differences in abuse estimates by interviewer.

Although sensitivity to violence can be worked on during interviewing training [22], it is not expected that the interviewer's personality changed. Therefore, we were expecting to find a relation with empathy, that is, because it would be easier to disclose personal history of abuse to a person with whom we feel more empathy. The relative

Table 3. Associations (OR 95%CI) between abuse and characteristics of participants and interviewers

Characteristics of participants and interviewers	Model 0	Model 1	Model 2
Gender (male)	—	0.84 (0.56–1.27)	0.84 (0.55–1.30)
Age (> 70 years)	—	0.99 (0.68–1.44)	1.02 (0.68–1.52)
Education (> 4 years)	—	0.89 (0.60–1.30)	0.94 (0.63–1.42)
Somatic complaints ^a (scoring > 19)	—	0.95 (0.62–1.44)	0.95 (0.61–1.49)
Anxiety ^b (scoring > 5)	—	1.36 (0.89–2.07)	1.30 (0.83–2.02)
Quality of life ^c (scoring > 89)	—	0.45 (0.30–0.67)	0.39 (0.26–0.60)
ECVC score	—	—	1.00 (0.95–1.05)
IRI score	—	—	1.01 (0.90–1.12)
ICC (%)	21.4	18.1	17.8

Abbreviations: ECVC, Escala de Crenças sobre Violência Conjugal; IRI, Interpersonal Reactivity Index; ICC, intraclass correlation coefficient; HADS, Hospital Anxiety and Depression Scale.

^a Giessen Complaint List.

^b HADS.

^c World Health Organization Quality of Life-old.

homogeneity of interviewers included in the present study regarding tolerance to violence and empathy scores may account for the observation that these factors did not explain differences in abuse estimates.

So far, we have discussed some points that could have been considered during study planning to minimize the interviewer variability. Although the occurrence of interviewer variability is well documented, little published information is available concerning its actual impact on inferences derived from multivariate data [7,23–25].

Our results showed substantial heterogeneity in estimates meaning that interviewers classify subjects differently with regard to certain sensitive topics in which the role of the interviewer–interviewee interaction is expected to gain importance. However, although differences in point estimates were observed, there was no statistical interaction, that is, the interviewer was not a major effect measure modifier. In support of this, the pooled estimates of associations in our study are concordant with previously described results regarding the associations between abuse and health status [26,27]. Therefore, we cannot discard the hypothesis that the observed variability in estimates of association between interviewers is random. Interviewer variability can theoretically be reduced by largely increasing the number of interviewers, but this scenario is seldom practical in epidemiologic research [28].

Even if effect heterogeneity between interviewers is found negligible, the existence of confounding by interviewer is simultaneously dependent on the types of exposure and outcome assessed. Indeed, when both the exposure and the outcome may be affected by the interviewer, such as in the association between somatic complaints or anxiety and abuse, the interviewer variable probably should be considered a confounder. There seemed to be little reason to believe that the interviewer has an effect on the report of variables such as age, and therefore, the interviewer is most probably not a confounder when assessing objective variables on abuse. Although, in our results, the point estimate of the exposure–outcome relationship was quite similar between crude and adjusted models,

it seems reasonable to consider the interviewer as a confounder when assessing the association between somatic complaints or anxiety and abuse.

Therefore, although no relevant effect modification was observed, this study emphasizes the importance of the interviewer as a relevant confounder when estimating associations between sensitive variables, particularly those related to social desirability, as it is the case of elder abuse.

Acknowledgments

The authors thank the ABUEL Group: Joaquim Soares (Sweden), Giovanni Lamura (Italy), Jutta Lindert (Germany), Elisabeth Ioannidi-Kapolou (Greece), Mindaugas Stankunas (Lithuania), and Francisco Torres-Gonzalez (Spain).

References

- [1] Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. *Lancet* 2002;360:1083–8.
- [2] Tolan P, Gorman-Smith D, Henry D. Family violence. *Annu Rev Psychol* 2006;57:557–83.
- [3] Clarke P, Sproston K, Thomas R. An investigation into expectation-led interviewer effects in health surveys. *Soc Sci Med* 2003;56:2221–8.
- [4] Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 4th ed. New York: Oxford University Press Inc.; 2008.
- [5] Bradburn NM, Sudman S. Interviewer variation in asking questions. In: Bradburn NM, Sudman S, editors. *Improving interview method and questionnaire design*. San Francisco, CA: Jossey-Bass Publishers; 1979:26–50.
- [6] Riessman CK. Interviewer effects in psychiatric epidemiology: a study of medical and lay interviewers and their impact on reported symptoms. *Am J Public Health* 1979;69:485–91.
- [7] Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. *Am J Epidemiol* 1997;146:429–38.
- [8] Davis RE, Couper MP, Janz NK, Caldwell CH, Resnicow K. Interviewer effects in public health surveys. *Health Educ Res* 2010;25(1):14–26.

- [9] Weiss CH. Interviewing in evaluation research. In: Struening EL, Guttentag M, editors. Handbook of evaluation research. Beverly Hills: Sage Publications; 1975:355–95.
- [10] Collyer CE, Brell A, Moster A, Furey J. Individual differences in sensitivity to violence. *Percept Mot Skills* 2011;113:703–14.
- [11] Soares J, Barros H, Torres-Gonzales F, Ioannidi-Kapolou E, Lamura G, Lindert J, et al. Abuse and health among elderly in Europe. Kaunas: Lithuanian University of Health Sciences Press; 2010.
- [12] Lindert J, Luna J, Torres-Gonzalez F, Barros H, Ioannidi-Kapolou E, Quattrini S, et al. Study design, sampling and assessment methods of the European study 'abuse of the elderly in the European region'. *Eur J Public Health* 2012;22(5):662–6.
- [13] Ramos E, Lopes C, Barros H. Investigating the effect of nonparticipation using a population-based case–control study on myocardial infarction. *Ann Epidemiol* 2004;14:437–41.
- [14] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- [15] Power M, Quinn K, Schmidt S, WHOQOL-OLD Group. Development of the WHOQOL-old module. *Qual Life Res* 2005;14: 2197–214.
- [16] Davis MH. A multidimensional approach to individual differences in empathy. *JSAS Catalog Selected Documents Psychol* 1980;10:85.
- [17] Machado C, Gonçalves M, Matos M. Escala de Crenças sobre Violência Conjugal e Inventário de Violência Conjugal. Escalas de avaliação e manual. Braga: Psiquilibrios; 2006.
- [18] Sampaio LR, Guimarães PRB, Camino CPS, Formiga NS, Menezes IG. Estudos sobre a dimensionalidade da empatia: tradução e adaptação do Interpersonal Reactivity Index (IRI). *Psico* 2011; 42(1):67–76.
- [19] Backstrom CH, Hursh-Cesar G. Survey research. 2nd ed. New York: Wiley; 1981.
- [20] Hornik J. Impact of pre-call request form and gender interaction on response to a mail survey. *J Marketing Res* 1982;19:144–51.
- [21] Pollner M. The effects of interviewer gender in mental health interviews. *J Nervous Ment Dis* 1998;186:369–73.
- [22] Collyer CE, Johnson KL, de Mesquita PB, Palazzo LA, Jordan D. Sensitivity to violence measured by ratings of severity increases after nonviolence training. *Percept Mot Skills* 2010;110(1):48–60.
- [23] Nybo Andersen AM, Olsen J. Do interviewers' health beliefs and habits modify responses to sensitive questions? A study using data collected from pregnant women by means of computer-assisted telephone interviews. *Am J Epidemiol* 2002;155:95–100.
- [24] Winters S, Strating MH, Klazinga NS, Kool RB, Huijsman R. Determining the interviewer effect on CQ Index outcomes: a multilevel approach. *BMC Med Res Methodol* 2010;10:75.
- [25] Davis P, Scott A. The effect of interviewer variance on domain comparisons. *Surv Methodol* 1995;21:99–106.
- [26] Cooper C, Selwood A, Livingston G. The prevalence of elder abuse and neglect: a systematic review. *Age Ageing* 2008;37:151–60.
- [27] Acierno R, Hernandez MA, Amstadter AB, Resnick HS, Steve K, Muzzy W, et al. Prevalence and correlates of emotional, physical, sexual, and financial abuse and potential neglect in the United States: the National Elder Mistreatment Study. *Am J Public Health* 2010; 100:292–7.
- [28] Edwards S, Slattery ML, Mori M, Berry TD, Caan BJ, Palmer P, et al. Objective system for interviewer performance evaluation for use in epidemiologic studies. *Am J Epidemiol* 1994;140: 1020–8.